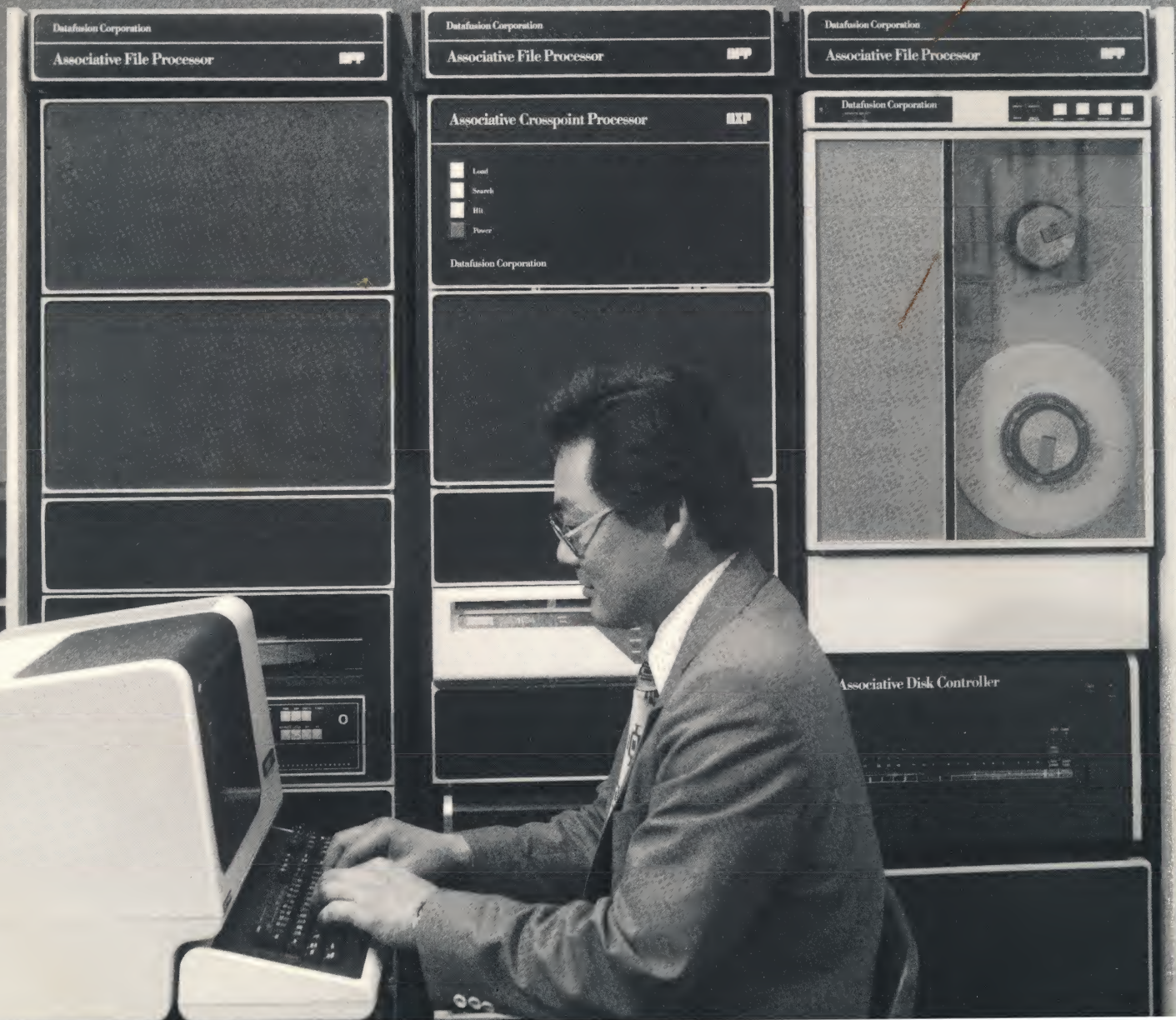




May 80

Datafusion Corporation

Associative File Processor AFP



The Associative File Processor AFP

A special purpose hardware system for retrieving textual information:

The automation of the office and of the publishing industry, the automation of message handling, the increased use of word processing and OCR devices, and the spread of digital communication networks are making enormous quantities of textual data available for computer processing. The conversion of textual information into digital formats will continue to grow at a phenomenal rate. Prior to the Associative File Processor, large text and document files could only be searched and retrieved by elaborate software systems, running on large main-frame computers. These software solutions to text retrieval have proven to be costly and laborious.

The Associative File Processor system is a low-cost, hardware based solution for text search and retrieval. The AFP® is truly a breakthrough in information processing and provides, for the first time, a simple, affordable, and rapid response alternative to conventional text processing software technology.

The AFP system moves the basic process involved in text retrieval out of software into a special purpose, parallel pipeline processor called the Associative Crosspoint Processor, or AXP.® The AXP performs the term matching function required in text searching, the most time consuming aspect of the information retrieval process. No other cpu resources are required for term matching. The AXP, together with conventional disk units, a PDP-11 minicomputer, control software, and a user front-end subsystem, constitute the Associative File Processor system.

Major Features of the AFP

Use Is Simple. English-like queries are used to search the AFP data base. Training to use the system takes only a few minutes and requires no special background to learn to operate.

Retrieval Is Simple. Documents or text material can be retrieved "on the fly" or batched. Perusal is either at display terminal or by hardcopy. Because the search is inexpensive, browsing the data base with general queries is cost-effective. Narrow, specific queries require only one pass through the data base.

Storage Is Simple. Convert source text to ASCII characters and write to any empty space on the search disk. No formatting, no indexing, no inversion.

Search Is Simple. Interactively enter queries. 8192 bytes of key word memory accommodates approximately 1200 key words. 40 to 70 complex queries processed simultaneously. Every word in data base is searched — no stop lists are needed. Handles boolean AND, OR, NOT and proximity key word logic. Query language allows Exact Match and "don't care" character (byte) logic.

Update Is Simple. The data base can be added to in seconds. New documents can be added to the system in between search cycles or "on the fly." Since the data is a collection of documents, the new material can be written onto any empty space. The entire document is preserved; no words are deleted.

Configuration Is Simple. A typical installation includes an AXP, disk units, a PDP-11, and occupies only one rack of equipment, plus the disk drives. No special room preparation such as raised floors or extra air conditioning is required. The system is off-the-shelf and available for delivery within 90 days.

Expandability Is Easy. The AFP system can be expanded by adding disks, by adding AXP's, and by adding displays. The system operates as a stand alone or as a subsystem to a large computer system.

Data Privacy Is Simple. The entire AFP system can fit in an average size, locked office. Removable data base disk packs can be locked in a safe, for added further privacy and security.

Purchase Price Is Low. Cost of the AFP is about the same as a typical data base management software package but has many orders of magnitude more processing power for text search and retrieval.

Operational Cost Is Low. Exceptional low cost. A few pennies per query, based on a reasonable utilization of the AFP system.



How the AFP System Works

The architecture of the AFP system is simple. Using conventional disk technology and Digital Equipment's PDP-11 minicomputer, the AXP compares the contents of the search disk at disk transfer rates up to 1.2 megabytes per second with 8192 bytes of parallel query term memory. The effective character matching at the maximum input rate is over 9 billion bytes per second.

A Typical AFP Configuration

PDP-11 Series CPU. Must have at least 96K words of memory.

System Disk and Controller. At least 5 megabytes. Recommend disk same size as Search Disk to hold user working files.

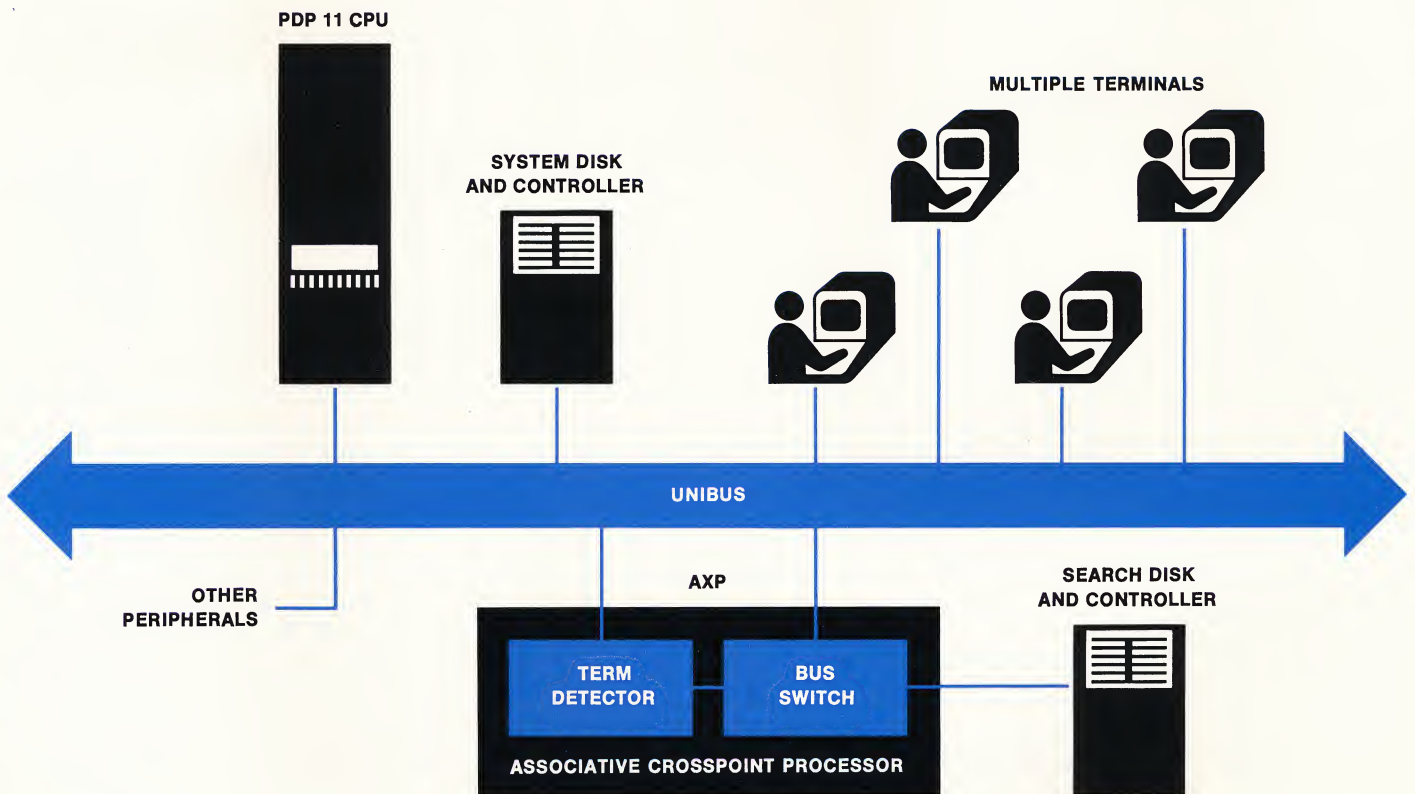
Multiple CRT Displays. Any TTY compatible display. Also compatible with TEMPEST tested displays, including UNIVAC 1652 and DEC VT-100.

The AXP. 8192 bytes of parallel query term memory, plus Bus Switch which disconnects Search Disk bus traffic from Unibus during search.

Search Disk and Controller. AXP will operate with a variety of disks and controllers including 300, 200 and 80 megabyte drives.

Other Peripherals. The AFP system supports magnetic tape, printer, removable disk systems and optical character readers. The AFP system can also support a communications front end for processing live message traffic.

AFP Associative File Processor Architecture



Six Basic Functions Are Supported by The AFP System

Data Base Maintenance. Data base loading, updating, and purging. In either on-line or batch mode.

Query Generation. On-line interactively with a simple query language.

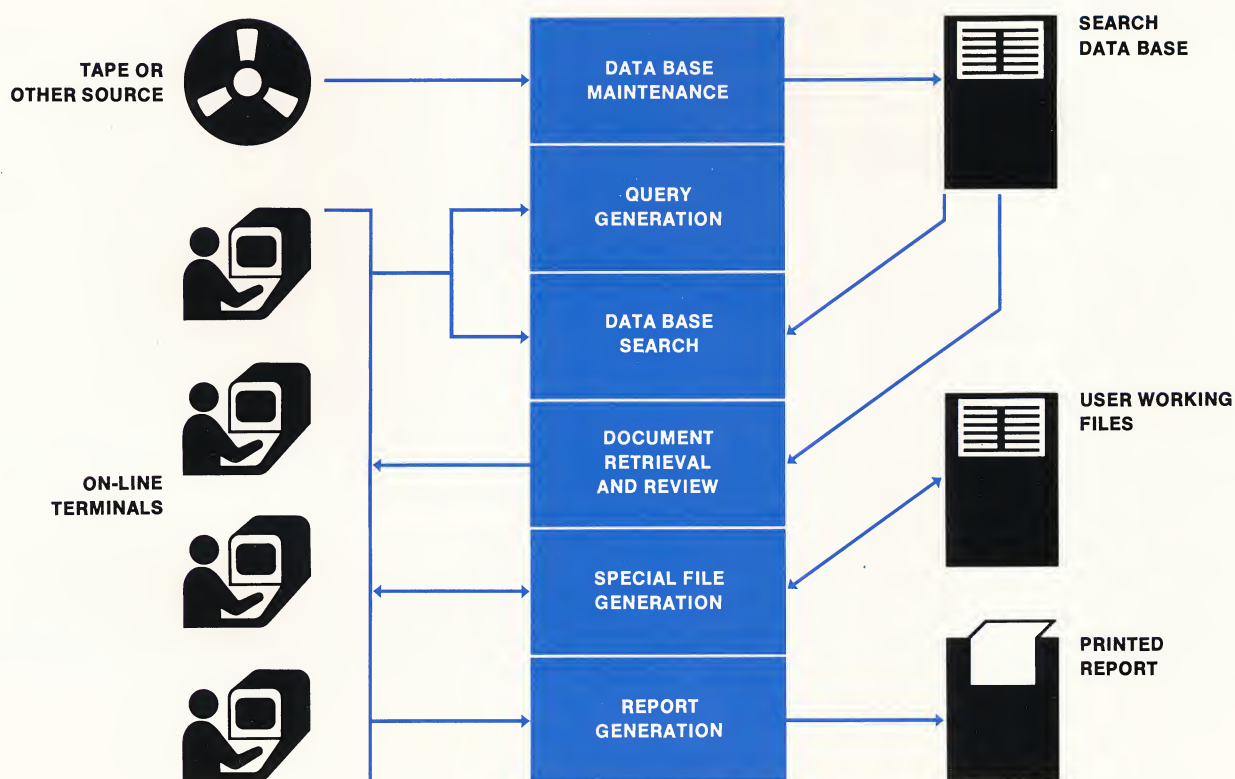
Data Base Search. Every word in the search data base is compared with the parallel query term memory.

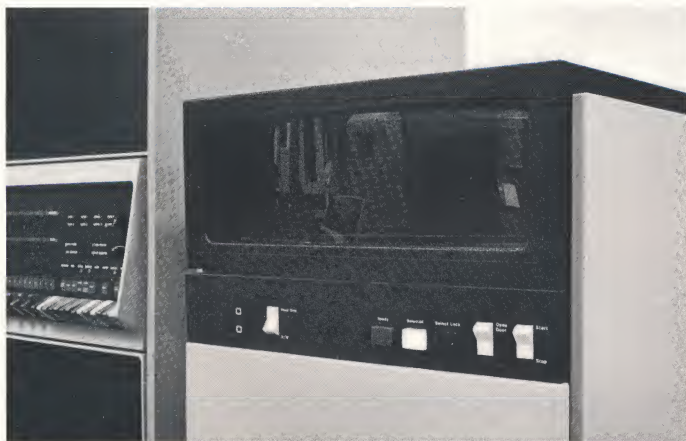
Document Retrieval and Review. Documents which satisfy a query for each user can be reviewed on CRT terminal and query terms can be highlighted on the screen.

Special File Generation. Documents of interest may be saved in a work file special to each user.

Report Generation. Any document may be printed in total or a special report abstracted from relevant documents may be composed and printed.

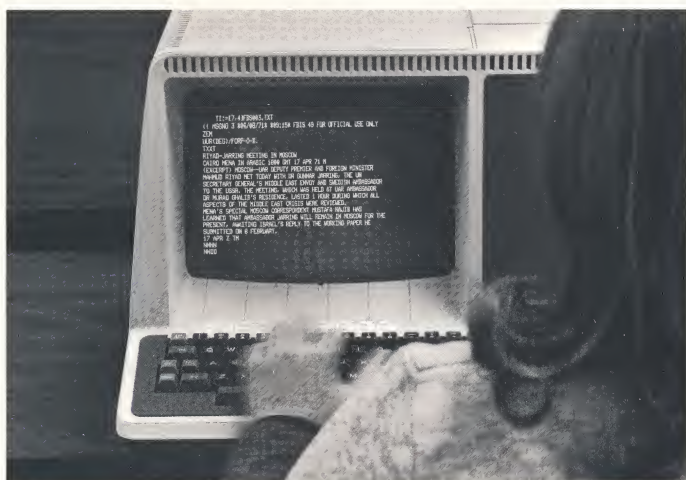
Functional Diagram of AFP





Data Base Maintenance. The AFP data base is normally maintained on 300 megabyte disk systems accessible by both the AFP hardware (the AXP) and the cpu. Initial data load is accomplished by transferring documents from source media onto the disk. Documents may be stored as completely free text or, if the source data is partially or fully formatted, this structure may be retained. A data load program, provided with the system, is used for building the data base. No special software and no indexing or inversion structures of any kind are necessary.

Subsequent file update is handled conventionally from the cpu and requires nothing more than adding, deleting, or revising the documents themselves.



Query Generation. Queries are generated on-line and may be input from several CRT display terminals simultaneously. All queries are collected and used in parallel during a data base search. The relationship between query generation and the search function is user controlled. A search may be performed with a single query or with dozens of queries from multiple terminals. The system resolves retrieval hits and delivers documents to the appropriate users. Queries can be entered in a natural language form or in a set expression form. Users can learn to enter queries and operate the AFP literally in a matter of minutes and typically become proficient and comfortable with the AFP after only a few search procedures.



Data Base Search. The search function takes place at user command. The AFP automatically collects all outstanding queries, extracts key terms and phrases, loads these into AXP hardware memory, and performs the search. For optimization, duplicate terms are loaded only once. Word-matches common to several queries are resolved by the query resolver. When the search function begins, the cpu bus is automatically disconnected from the search disk and a continuous stream of search data is read into the AXP. The stream of text words for each document is matched against the query terms by the AXP hardware logic. During a search the cpu is free for other chores since its bus is disconnected from the search disk. No software is involved in the matching process.

As key terms are found, the AFP determines whether they satisfy any of the outstanding queries and if so, documents are added to a hit list for the appropriate user. The search disk can contain a single search file or many

search files can coexist on one search disk. A single search file occupying an entire 300 million byte disk can be searched in 4 - 5 minutes. A search file less than 300 million bytes requires proportionally less time to search.

Document Retrieval and Post Processing. The retrieval and review of hit documents is performed on user demand and can be done either during or after a search. If immediate review is required, documents are displayed as soon as they are found, interrupting the search to do so. In this case, review can commence within seconds of initiating the search. In either case, document hit lists are preserved and can be manipulated in a variety of ways. Documents may be reviewed on a CRT display. Page and document skipping forward and backward is provided. A feature to highlight matched query terms in the document is provided at the option of the user. This feature is extremely useful in making relevant passages in the document relate to the query. Documents of interest may be saved in a user's personal file for review at a later time or may be output to other devices such as a magnetic tape unit, a printer, or sent over a communication line. A report generation capability is available allowing the user to abstract relevant passages from documents of interest and to compose a report which may be output to various media.

Available AFP Configurations. The AFP can be delivered in a variety of different configurations:

AXP-100. This product consists of the Associative Cross-point Processor (AXP), the firmware and the user interface software. It is attachable to any existing PDP11 series computer at the customers site provided it is of adequate configuration. It can operate under RSX-11M, RSX-11D or IAS operating systems. THE AXP-100 can be supplied with an 80 megabyte (as an AXP-100D80) or a 300 megabyte (as an AXP-100D300) associative disk and controller.

AXP-200. This product is an Associative File Processor subsystem attachable to virtually any other computer type such as an IBM, Univac, a CDC, a Data General, etc. The interface to the other computer is through a standard RS232 interface. The AXP-200 consists of the AXP-100D300, a PDP11 CPU, a system disk and controller, and a RS232 interface. The user terminals are part of the customer's existing computer. The AXP-200 takes full control of the search and retrieval leaving the customers computer completely independent and capable of running other programs.

AXP-300. This product is a turnkey, stand-alone Associative File Processor configuration supplied with an AXP-100D300, a PDP11 computer a system disk and controller, a high speed printer, a magnetic tape drive, four CRT terminals and operating software.

Specials. Special configurations and custom software will be quoted on request.



How to Use the AFP

There Are Two Forms For Entering Queries

Queries are unrestricted in vocabulary and can be entered in the two forms shown on the right. The natural language format allows a new user unfamiliar with structuring queries to initially use the AFP system with minimal instruction. Any natural language query sentence can be composed so long as single quotation marks are placed around a key word or key contiguous word phrase of interest. This allows a new AFP user to get started instantly. The interactive language provides a learning environment for the user, since useful information is retrieved from the search immediately. Near real time response encourages the user to learn how to construct better and more efficient queries. On the right an efficient query is composed by using the boolean expression form.

Natural Language Form for Entering a Query

FIND ALL RESEARCH ON 'VEHICLE BRAKING' PARAMETERS AND EFFECTS OF 'LOAD SENSING VALVES' AND IN REDUCTION OF 'TRACTOR TRAILER' 'JACKKNIFING' BUT DO NOT INCLUDE 'TIRE' PARAMETERS.

Boolean Expression Form for Entering a Query

'VEHICLE BRAKING' AND 'LOAD SENSING VALVES' AND 'TRACTOR TRAILER' AND 'JACKKNIFING' NOT 'TIRE'

Synonyms in Queries

Better queries can usually be constructed by including synonyms of either key words or key contiguous word phrases. Query synonyms are accomplished by using the logical OR operator. For instance, the query example on the right above in the boolean expression form could be expanded using synonyms with the OR operator as shown opposite.

Boolean Expression Form With Synonyms

'VEHICLE BRAKING' OR 'AIR BRAKES' OR 'TRUCK BRAKING' AND 'LOAD SENSING VALVES' OR 'WEIGHT SENSING' AND 'TRACTOR TRAILER' OR 'TRUCK' OR 'TRUCKER' AND 'JACKKNIFING' OR 'CRASH' OR 'ACCIDENT' NOT 'TIRE' OR 'TIRES'

Word Proximity in Queries

Proximity searching is used to delimit the context or the discourse surrounding a word or group of words. The proximity feature allows a query to be more precise and be used to identify phrases, sentences, and paragraph boundaries. The example on the right would find all documents in which "Tractor Trailer" and "Jackknifing" occurred providing that "Jackknifing" followed "Tractor Trailer" and no more than 18 text words intervened between "Trailer" and "Jackknifing."

Boolean Expression Form With Word Proximity

'TRACTOR TRAILER' 20 'JACKKNIFING'

Numeric Ranging in Queries

Numeric ranging is accomplished by using the Don't Care Character "@" and by enumerating the possible intervening number combinations. An example is shown at right.

An Example of Numeric Ranging

Find all ages between 29 and 50:
'AGE' AND '29' OR '3@' OR '4@' OR '50'

Don't Care Characters in Queries

Don't Care Characters are represented by a "@" character and are useful for a number of situations, including spelling differences, suspected spelling errors, typing character transposition errors, affix and suffix stripping, (e.g. to retrieve chemical compounds). Don't Care Characters may be leading, trailing or intermediate characters, in any combination.

Some Examples of Don't Care Characters

Spelling Differences	Replace With	'KHRUSHCHEV' OR 'KHRUSHCHOV' OR 'KHRUSHCH@V'
Suffix Stripping	Replace With	'RIOT' OR 'RIOTS' OR 'RIOTED' OR 'RIOTING' OR 'RIOT@' OR 'RIOT@@" OR 'RIOT@@@'
Affix Stripping	Replace With	'S2102' OR 'EA2102' OR 'TMS2102' OR '@2102' OR '@@2102' OR '@@@2102' (thus matching also all other manufacturers of 1K RAMs)
Chemical Terms	Replace With	'SULF . . . various endings' (more than fifty words) OR 'SULF@' OR 'SULF@@" OR 'SULF@@@' OR 'SULF@@@@' OR . . . 'SULF (longest stem)'

Don't Care Words Within Contiguous Word Phrases in Queries

Don't Care Words are represented by "*" character in a query and are used in contiguous word phrases. Two examples are shown opposite involving an unknown middle name or initial or a persons name and a date range. It is also used for representing text words in applications where there are known variations in the phrase sought.

Some Examples of Don't Care Words

Unknown Middle Name or Initial	Replace With	'JOHN A SMITH' OR 'JOHN ARNOLD SMITH' OR . . . 'JOHN Z SMITH' OR 'JOHN * SMITH'
Date Range	Replace With	'JANUARY 1 1979' OR 'JANUARY 2 1979' OR . . . 'JANUARY 31 1979' OR 'JANUARY * 1979'

Software Aids in Selecting and Using Synonyms

The AFP software has several synonym related aids. The first is a program which generates a concordance of the uncommon words in the search file and makes it available in hard copy form. By reviewing the concordance, several synonyms are usually obvious to the user and assists him in selecting synonyms. The second synonym aid allows the user to build his own personal synonym thesaurus. A synonym group can be specified by the user by giving it a name and placing it in his own thesaurus file. During query input the user need only specify the synonym name and all of the words in the synonym group will be automatically entered into the query. The third synonym aid allows for a synonym thesaurus common and accessible to all users. Thus by specifying a synonym name directed to the common thesaurus, all of the words in that synonym group can be automatically entered into the query.

Canned Queries

Queries used over and over again can be given a name and stored in a user file. By calling that name the entire query can be retrieved, presented on the CRT screen, and used. It may also be edited and modified before entering the query for searching.

Software Aid For Statistical Analysis

User application programs may access specific data fields from a virtual document hit list. Under user software control data may be extracted and organized for statistical analysis.

Associative File Processor Applications

The AFP for the first time provides an affordable, easy to use alternative for the search and retrieval of large textual data bases. These data bases can exist in many mixed forms, such as messages, memos, letters, reports, articles, books, documents and including formatted data forms.

Functional Applications Include

Data Input Media

Word Processing Data Entry
Optical Character Reader
Live Communication Line

Data Output Media

Hardcopy Reports
Routing by Communication Line
Removable Disk or Tape

Search Data Types

Textual Files
Formatted Files
Mixed Text and Formatted Files

Data Security

Controlled Locked AFP Room
Removable Search Disk Files
Removable Personal Files

Application Areas Include

Military and Intelligence

Law Enforcement

Library Search

Word Processing Support

Title and Property Search

Product Bibliographies

Litigation Support

Chemical Compound Retrieval

Historical Records and Archives

Technical Report Retrieval

Generic Record Keeping

Current Awareness Bulletin

Abstract Search

Laboratory Testing and Retrieval

Journal Abstracting and Control

Trial Transcripts

Pharmaceutical Literature Retrieval

Patent Search

The AFP system can also be expanded or configured differently to support other more pattern-related search and retrieval applications. We look forward to the opportunity of discussing your particular application with you.





ASSOCIATIVE FILE PROCESSOR (AFP)

PRICE LIST

	<u>Price*</u>	<u>Installation</u>
AXP-100 Includes Associative Crosspoint Processor, the firmware and user software with software license and documentation (1 set)	\$ 58,000	\$ 2,420
AXP-100D300 Includes AXP-100, plus 300 megabyte disk drive, disk controller and one disk pack	\$ 82,500	\$ 3,170
AXP-100D80 Includes AXP-100, plus 300 megabyte disk drive, disk controller and one disk pack	\$ 77,000	\$ 3,020
AXP-200 Includes AXP-100D300, a PDP 11 CPU, a system disk and controller, a communication interface and one set of documentation. A RS232 interface is standard. A high speed DMA interface is available for; 1) IBM 370 multiplexer, block multiplexer, or selector channel, 2) Univac 494, 1108, 1110 or 1143 processors, 3) CDC 6600 or 7600 processors, 4) Burroughs B6700 or B7700 processors, 5) Data General 300, 6) U.S. Navy NTDS, 7) Honeywell 6000, and 8) ARPANET IMP	\$130,000 to \$150,000 (depending on inter- face)	3% of total price
AXP-300 Includes AXP-100D300, a PDP 11 CPU, a system disk and controller, an 800/1600 BPI 9 track magnetic tape unit, a high speed line printer, four CRT terminals and one set of documentation	\$180,000 to \$200,000 (typical depending on peri- pheral selection)	3% of total price

Quantity and OEM discounts are available

*Prices are FOB Los Angeles. They are subject to change without notice.

APR 1980



Datafusion Corporation

21031 Ventura Boulevard, Woodland Hills, CA 91364 Phone (213) 887-9523

5 May 1980

Mr. T. Nelson
Box 3
Schooley Mountain, NJ 07870

Thank you for your interest in the "Associative File Processor" (AFP).

Enclosed is a brochure that will more fully explain the AFP and its potential uses.

Please contact me for any additional information that you may require.

Sincerely,

DATAFUSION CORPORATION

Jack W. Aubuchon
Marketing Representative

Enclosure

AFP Specifications

CPU	PDP 11/23, 34, 35, 40, 44, 45, 60, 70, or VAX with 96K words minimum.
Maximum Numbers of Queries	8192 bytes of query terms or 40-70 complex queries containing multiple synonyms, boolean operators, proximity logic and don't care characters
Retrieval Response	Typically a few seconds. A document is retrieved at time of query match.
Maximum Search Time	4 to 5 minutes with 200 or 300 megabyte disk
CRT Terminals	Standard of 4, expandable to 48.
Reliability	6400 hours MTBF for AXP
Power	115VAC, 60Hz, 4 amps single phase for AXP and Busrouter 115VAC, 60Hz, 6 amps single phase for disk controller if supplied 220VAC, 60Hz, 27 amps three phase for disk drive if supplied Additional power for PDP 11 system if supplied
Physical	One standard DEC cabinet
Location	AXP cabinet must be no more than 2 cabinet spaces from CPU

The Associative File Processor is protected by patents issued by the U.S. Patent Office.

DEC, PDP-11, VAX, UNIBUS are registered trademarks of Digital Equipment Corporation.

For more information or for a demonstration of the Associative File Processor call or contact a Marketing Representative at Datafusion Corporation.



Head Office

Datafusion Corporation
21031 Ventura Boulevard
Woodland Hills, California 91364
Telephone: (213) 887-9523
Attention: Marketing Representative

East Coast

Datafusion Corporation
Post Office Box 2909
Reston, Virginia 22090
Telephone: (703) 860-0022
Attention: Joe Caplan

